Universal coding for transmission of private information

Nilanjana Datta[1, a)] and Min-Hsiu Hsieh[2, b)]

[1)] *Statistical Laboratory, University of Cambridge, Wilberforce Road, Cambridge CB3 0WB, UK*

[2)] *ERATO-SORST Quantum Computation and Information Project, Japan Science and Technology Agency, 5-28-3, Hongo, Bunkyo-ku, Tokyo, Japan 113-0033*

We consider the scenario in which Alice transmits private classical messages to Bob via a classical-quantum channel, part of whose output is intercepted by an eavesdropper Eve. We prove the existence of a universal coding scheme under which Alice's messages can be inferred correctly by Bob, and yet Eve learns nothing about them. The code is universal in the sense that it does not depend on specific knowledge of the channel. Prior knowledge of the probability distribution on the input alphabet of the channel, and bounds on the corresponding Holevo quantities of the output ensembles at Bob's and Eve's end suffice.

arXiv:1007.2629v3 [quant-ph] 1 Jan 2011

[a)] Electronic mail: N.Datta@statslab.com.ac.uk

[b)] Electronic mail: minhsiuh@gmail.com

I. INTRODUCTION

A quantum channel can be used for a variety of different purposes and, unlike classical channels, it has many different capacities depending on what it is being used for, on the nature of its inputs and what additional resources are available to the sender and the receiver. In addition to its use in conveying classical and quantum information, and generating entanglement, a quantum channel can also be used to convey *private* classical information which is inaccessible to an eavesdropper. This allows unconditionally secure key distribution, which is impossible in the classical realm.

The different capacities of a quantum channel were first evaluated under the assumption that the channel was memoryless, that is, correlations in the noise acting on successive inputs to the channel were assumed to be absent. Holevo[1] and Schumacher and Westmoreland[2] proved that the classical capacity of a memoryless quantum channel under the restriction of product-state inputs, is given by the so-called Holevo capacity, the unrestricted classical capacity then being obtained by a regularisation of this quantity. An expression for the private classical capacity was independently obtained by Cai *et al.*[3] and by Devetak[4], who is also credited with the first rigorous proof of the expression for the quantum capacity (first suggested by Lloyd[5] and further justified by Shor[6]).

An inherent assumption underlying all these results is that the quantum channel is known *perfectly* to Alice and Bob. This assumption is, however, not necessarily valid in real-world communication systems, since it might be practically impossible to determine all the parameters governing a quantum channel with infinite accuracy. Thus one often has only limited knowledge of the quantum channel which is being used. This calls for the design of more general communication protocols which could be used for transmission of information through a quantum channel in spite of such a channel uncertainty. The corresponding coding theorems are then *universal* in the sense that they do not rely on exact knowledge of the channel used.

In the quantum setting, progress in this direction was first made by Datta and Dorlas[7] who obtained an expression for the classical capacity of a convex combination of memoryless quantum channels. This corresponds to the case in which Alice and Bob's only prior knowledge is that the channel in use is one of a given finite set of memoryless channels, with a given prior probability. It is hence the simplest model with channel uncertainty. This result

was further generalized and extended by Bjelakovic *et al.*[8,9] who derived the classical and quantum capacities of the so-called *compound quantum channels*, in which the underlying set of memoryless channels was allowed to be countably infinite or even uncountable. They also evaluated the optimal rates of entanglement transmission and entanglement generation through such channels[10]. In the classical setting, the first study of channel uncertainty dates back to the work of Wolfowitz[11,12], and of Blackwell *et al*[13], who determined the capacity of compound classical channels.

Note that when Alice is interested in sending only classical messages through a quantum channel Φ, she first needs to encode her message into a state of a quantum system which can then be transmitted through the channel. Denoting this encoding map by $\mathcal{E}$, one effectively obtains a *classical-quantum* $(c \to q)$ channel $W := \Phi \circ \mathcal{E}$ which maps classical messages into quantum states in the output Hilbert space of the channel Φ. Hayashi[14] proved a universal coding theorem for memoryless $c \to q$ channels, as a quantum version of the classical universal coding by Csiszár and Körner[15].

In this paper, we consider transmission of *private* classical information through a $c \to qq$ channel from Alice to Bob and Eve, and prove the existence of a universal code, for which the private capacity of the channel is an achievable rate. The channel is defined by the map $W : x \longrightarrow W^{BE}(x)$ with $x \in \mathcal{X}$ (a finite classical alphabet) and $W^{BE}(x)$ being a state defined on a bipartite quantum system BE. Bob has access to the subsystem B, whereas Eve (the eavesdropper) has access to the subsystem E. Such a channel induces two $c \to q$ channels – one from Alice to Bob (which we denote by W^B), and one from Alice to Eve (which we denote by W^E). We prove a universal private coding scheme under which Bob can infer Alice's message with arbitrary precision in the asymptotic limit, simultaneously ensuring that Eve learns arbitrarily little about the message. The code is universal in the sense that it does not depend on knowledge of the structure of the channel W^{BE}. The only assumption in the coding theorem is that Alice and Bob have prior knowledge of the input distribution $\underline{p}$ on the set $\mathcal{X}$, and of bounds on the corresponding Holevo quantities for the channels W^B and W^E.

As a first step towards proving a universal private coding theorem, we derive an alternative proof of a universal coding theorem for a memoryless $c \to q$ channel (see Theorem 1 of Section III). Our coding theorem for the $c \to q$ channel W^B, only requires prior knowledge of the probability distribution on the input of the channel. It establishes the existence of a universal

3

code using which Alice and Bob can achieve reliable information transmission through the $c \to q$ channel at any rate less than the corresponding Holevo quantity. Our proof employs a "type decomposition" and the random coding technique, but unlike Hayashi's proof[14], it does not employ irreducible representations and Schur-Weyl duality. However, the decoding POVM in our universal code is analogous to his, which results in some of the steps of our proof being similar. Moreover, like his result, our theorem can be essentially viewed as a universal version of a cornerstone of information theory, namely, the packing lemma[15,16].

The universal packing lemma ensures that Bob correctly infers Alice's messages in the asymptotic limit. In addition, we require that these messages cannot be inferred by the eavesdropper, Eve. This obliteration of information transmitted over the channel W^E, induced between Alice and Eve, is established by employing the so-called covering lemma[17,18], which we prove explicitly below. We also establish that the covering lemma is universal because it does not require Alice to have any specific knowledge of the channel W^E. It only depends on the Holevo quantity corresponding to the input distribution of the channel. The universal covering lemma, when combined with the universal packing, yields our main result, namely, the universal private coding theorem.

Universal coding theorems have been established for other information-processing tasks, for example, data compression[19–21] and entanglement concentration[22,23]. Jozsa *et al*[19] introduced a universal data compression scheme which did not require any knowledge about the information source, other than an upper bound on its von Neumann entropy. The compression scheme was proved to achieve a rate equal to this upper bound. Hence, our first result, Theorem 1 of Section III, can be viewed as the $c \to q$ channel counterpart of this result. Similarly, our second (and main) result, Theorem 3 of Section V, is in a way the $c \to qq$ channel counterpart of this same result, under the additional requirement of privacy.

Note that the work by Jozsa *et al*[19] was followed by fully universal quantum data compression schemes, presented first by Hayashi and Matsumoto[21] and then by Jozsa and Presnell[20], in which the von Neumann entropy of the quantum information source was not known apriori but was instead estimated.

In Section II we introduce the relevant notations and definitions. In Section III we prove a universal coding theorem for a $c \to q$ channel. The Universal covering lemma is proved in Section IV. In Section V, the results of the previous sections are combined to prove our main result, namely a universal private coding theorem. We conclude in Section VI.

II. NOTATIONS AND DEFINITIONS

Let $\mathcal{B}(\mathcal{H})$ denote the algebra of linear operators acting on a finite–dimensional Hilbert space $\mathcal{H}$ and let $\mathfrak{D}(\mathcal{H})$ denote the set of positive operators of unit trace (states) acting on $\mathcal{H}$. We denote the identity operator in $\mathcal{B}(\mathcal{H})$ by $\mathbb{I}$. For a state $\rho \in \mathfrak{D}(\mathcal{H})$, the von Neumann entropy is defined as $S(\rho) := -\operatorname{Tr}\rho\log\rho$. Further, for a state ρ and a positive operator σ such that $\operatorname{supp}\rho \subseteq \operatorname{supp}\sigma$, the quantum relative entropy is defined as $S(\rho||\sigma) = \operatorname{Tr}\rho\log\rho - \rho\log\sigma$, whereas the relative Rényi entropy of order $\alpha \in (0,1)$ is defined as

$$S_\alpha(\rho||\sigma) := \frac{1}{\alpha-1}\log\big[\operatorname{Tr}(\rho^\alpha\sigma^{1-\alpha})\big].$$

Two entropic quantities, defined for any ensemble of states $\mathcal{E} := \{p_x, \sigma_x\}_{x\in\mathcal{X}}$, play a pivotal role in this paper. One is the Holevo quantity, which is given in terms of the quantum relative entropy as follows:

$$\begin{aligned}
\chi(\mathcal{E}) &= \min_{\omega_Q} S(\sigma_{XQ}||\sigma_X \otimes \omega_Q) \\
&= S(\sigma_{XQ}||\sigma_X \otimes \sigma_Q) \\
&= S\Big(\sum_x p_x\sigma_x\Big) - \sum_x p_x S(\sigma_x),
\end{aligned} \tag{1}$$

where σ_{XQ} is a classical-quantum state

$$\sigma_{XQ} := \sum_x p_x |x\rangle\langle x| \otimes \sigma_x,$$

and σ_X, σ_Q denote the corresponding reduced states. The second identity in (1) follows from the fact[24]:

$$\min_{\omega_Q} S(\sigma_{XQ}||\sigma_X \otimes \omega_Q) = S(\sigma_{XQ}||\sigma_X \otimes \sigma_Q). \tag{2}$$

The other relevant entropic quantity is the α-χ quantity, which is defined for any $\alpha \in (0,1)$ in terms of the relative Rényi entropy of order α as follows:

$$\begin{aligned}
\chi_\alpha(\mathcal{E}) &:= \min_{\omega_Q} S_\alpha(\sigma_{XQ}||\sigma_X \otimes \omega_Q) \\
&= \frac{\alpha}{\alpha-1}\log\operatorname{Tr}\left[\sum_{x\in\mathcal{X}} p_x\sigma(x)^\alpha\right]^{\frac{1}{\alpha}},
\end{aligned} \tag{3}$$

(For a proof of the last identity, see e.g.[25]).

It is known that (see e.g.[26])

$$\lim_{\alpha\nearrow 1} S_\alpha(\rho||\sigma) = S(\rho||\sigma).$$

Moreover, it has been proved (see Lemma B.3[27]) that

$$\lim_{\alpha \nearrow 1} \chi_\alpha(\mathcal{E}) = \chi(\mathcal{E}). \tag{4}$$

Throughout this paper we take the logarithm to base 2 and restrict our considerations to finite-dimensional Hilbert spaces.

The trace distance between two operators A and B is given by

$$\|A - B\|_1 := \text{Tr}\big[\{A \geq B\}(A - B)\big] - \text{Tr}\big[\{A < B\}(A - B)\big],$$

where $\{A \geq B\}$ denotes the projector onto the subspace where the operator $(A - B)$ is non-negative. We make use of the following lemmas:

Lemma 1.[28] *Given a state ρ and a self-adjoint operator ω, for any real γ we have*

$$\text{Tr}\big[\{\rho \geq 2^{-\gamma}\omega\}\omega\big] \leq 2^\gamma.$$

Lemma 2 (Gentle measurement lemma[29,30]). *For a state $\rho \in \mathfrak{D}(\mathcal{H})$ and operator $0 \leq \Lambda \leq \mathbb{I}$, if $\text{Tr}(\rho\Lambda) \geq 1 - \delta$, then*

$$\left\|\rho - \sqrt{\Lambda}\rho\sqrt{\Lambda}\right\|_1 \leq 2\sqrt{\delta}.$$

The same holds if ρ is a subnormalized density operator.

Lemma 3 (Operator Chernoff bound[31]). *Let $\sigma_1, \cdots \sigma_N$ be independent and identically distributed random variables with values in $\mathcal{B}(\mathcal{H})$, which are bounded between 0 and the identity operator $\mathbb{I}$. Assume that the expectation value $\mathbb{E}\sigma_m = \Omega \geq t\mathbb{I}$ for some $0 < t < 1$. Then for every $0 < \varepsilon < 1/2$*

$$\Pr\left\{\frac{1}{N}\sum_{m=1}^{N}\sigma_m \notin [1 \pm \varepsilon]\Omega\right\} \leq 2\dim\mathcal{H}2^{\left(-Nk\varepsilon^2 t\right)} \tag{5}$$

where $k := 1/(2(\ln 2)^2)$, and $[1 \pm \varepsilon]\theta = [(1-\varepsilon)\theta; (1+\varepsilon)\theta]$ is an interval in the operator order: $[A; B] = \{\sigma \in \mathcal{B}(\mathcal{H}) : A \leq \sigma \leq B\}$.

Lemma 4.[14] *For any operator $A \geq 0$ and $t \in (0, 1)$, we have*

$$\max_{\sigma \in \mathfrak{D}(\mathcal{H})} \text{Tr}\left(A\sigma^t\right) = \left[\text{Tr}\left(A^{\frac{1}{1-t}}\right)\right]^{1-t}.$$

6

We state here a number of standard facts about types and typical sequences[15,32], which we use in this paper. Let $\mathcal{X}$ denote a finite classical alphabet of size $|\mathcal{X}| = k$, and let the letters of the alphabet $\mathcal{X}$ be ordered (e.g. lexicographically):

$$\mathcal{X} := \{\mathsf{x}_1, \mathsf{x}_2, \ldots, \mathsf{x}_k; \mathsf{x}_1 \leq \mathsf{x}_2 \leq \ldots \leq \mathsf{x}_k\}.$$

Let us denote by $N(\mathsf{x}_i | x^n)$ the number of occurrences of the symbol $\mathsf{x}_i \in \mathcal{X}$ in the sequence $x^n := (x_1, \cdots, x_n) \in \mathcal{X}^n$. We define the *type* $t(x^n)$ of a sequence $x^n \in \mathcal{X}^n$ as follows: $t(x^n) := \underline{q}$, where $\underline{q} = (q_1, q_2, \ldots, q_k)$, denotes a probability vector of length k with elements

$$q_i = \frac{N(\mathsf{x}_i | x^n)}{n}.$$

Let $\mathcal{P}_{\mathcal{X}}^n$ denote the set of types in $\mathcal{X}^n$. The size of $\mathcal{P}_{\mathcal{X}}^n$ is bounded as follows[32]:

$$|\mathcal{P}_{\mathcal{X}}^n| \leq (n+1)^k = 2^{n\zeta_n(k)}, \tag{6}$$

where, for any fixed integer k, we define

$$\zeta_n(k) := \frac{k}{n} \log(n+1). \tag{7}$$

Note that $\zeta_n(k) \to 0$ as $n \to \infty$. Define the set of sequences of type $\underline{q}$ in $\mathcal{X}^n$ by

$$\mathcal{T}_{\mathcal{X}}^n(\underline{q}) = \{x^n \in \mathcal{X}^n : t(x^n) = \underline{q}\}.$$

For any type $\underline{q} \in \mathcal{P}_{\mathcal{X}}^n$, we have[32]:

$$2^{n[H(\underline{q}) - \zeta_n(k)]} \leq |\mathcal{T}_{\mathcal{X}}^n(\underline{q})| \leq 2^{nH(\underline{q})}. \tag{8}$$

Throughout this article, we denote the probability distribution on the set $\mathcal{X}$ by $\underline{p}$, i.e., $\underline{p} = (p_1, p_2, \cdots, p_k)$, where $p_i := p_{\mathsf{x}_i}$ for $\mathsf{x}_i \in \mathcal{X}$. The Shannon entropy of $\underline{p}$ is defined as $H(\underline{p}) = -\sum_{i=1}^{k} p_i \log p_i$. For any $\delta > 0$, define $\mathcal{P}_{\underline{p},\delta}^n := \{\underline{q} \in \mathcal{P}_{\mathcal{X}}^n : |q_i - p_i| \leq p_i \delta, \ \forall \mathsf{x}_i \in \mathcal{X}\}$. Define the set of δ-typical sequences of length n as

$$\mathcal{T}_{\underline{p},\delta}^n = \bigcup_{\underline{q} \in \mathcal{P}_{\underline{p},\delta}^n} \mathcal{T}_{\mathcal{X}}^n(\underline{q})$$

$$= \left\{ x^n \in \mathcal{X}^n : \left| \frac{N(\mathsf{x}_i | x^n)}{n} - p_i \right| \leq p_i \delta, \forall \mathsf{x}_i \in \mathcal{X} \right\}. \tag{9}$$

For any $\varepsilon, \delta > 0$, some positive constant $c := H(\underline{p})$ depending only on $\underline{p}$, and sufficiently large n, we have[32]:

$$Q_n := \Pr\{X^n \in \mathcal{T}^n_{\underline{p},\delta}\} \geq 1 - \varepsilon \tag{10}$$

$$2^{-n[H(\underline{p})+c\delta]} \leq p^n_{x^n} \leq 2^{-n[H(\underline{p})-c\delta]}, \forall x^n \in \mathcal{T}^n_{\underline{p},\delta} \tag{11}$$

$$|\mathcal{T}^n_{\underline{p},\delta}| \leq 2^{n[H(\underline{p})+c\delta]}, \tag{12}$$

where $p^n_{x^n}$ denotes the probability of the sequence x^n and is given by the product distribution $p^n_{x^n} := \prod^n_{i=1} p_{x_i}$. We also use the following bound[32]: For any type $\underline{q} \in \mathcal{P}^n_{\underline{p},\delta}$,

$$|\mathcal{T}^n_{\mathcal{X}}(\underline{q})| \geq 2^{n\left[H(\underline{p})-\eta(\delta)\right]}, \tag{13}$$

where $\eta(\delta) \to 0$ as $\delta \to 0$.

Consider a Hilbert space $\mathcal{H}$, where $\dim \mathcal{H} = d$. Let $\mathcal{Y} = \{1, 2 \cdots, d\}$. It follows that $\mathcal{H}^{\otimes n} = \mathrm{Span}\{|y^n\rangle \equiv |y_1\rangle \otimes \cdots \otimes |y_n\rangle : \forall y^n \in \mathcal{Y}^n\}$. Let $\mathcal{K}_{\underline{q}} = \mathrm{Span}\{|y^n\rangle : y^n \in \mathcal{T}^n_{\mathcal{Y}}(\underline{q})\}$, where $\mathcal{T}^n_{\mathcal{Y}}(\underline{q})$ is the collection of sequences of type $\underline{q}$ in $\mathcal{Y}^n$. Then

$$\mathcal{H}^{\otimes n} = \bigoplus_{\underline{q} \in \mathcal{P}^n_{\mathcal{Y}}} \mathcal{K}_{\underline{q}},$$

where $\mathcal{P}^n_{\mathcal{Y}}$ is the collection of all types in $\mathcal{Y}^n$. Let $\widetilde{\mathcal{K}}_{\underline{q}} = \mathrm{Span}\{U^{\otimes n}|y^n\rangle : \forall U \in \mathrm{U}(d), y^n \in \mathcal{T}^n_{\mathcal{Y}}(\underline{q})\}$, where $\mathrm{U}(d)$ is the group of $d \times d$ unitary matrices. Note that $\widetilde{\mathcal{K}}_{\underline{q}}$ is not associated with a preferred basis, unlike $\mathcal{K}_{\underline{q}}$. Let $I_{\underline{q}} \in \mathcal{B}(\mathcal{H}^{\otimes n})$ be the projector onto $\mathcal{K}_{\underline{q}}$:

$$I_{\underline{q}} = \sum_{y^n \in \mathcal{T}^n_{\mathcal{Y}}(\underline{q})} |y^n\rangle\langle y^n|, \tag{14}$$

and let $\widetilde{I}_{\underline{q}} \in \mathcal{B}(\mathcal{H}^{\otimes n})$ be the projector onto $\widetilde{\mathcal{K}}_{\underline{q}}$. Since $\mathcal{K}_{\underline{q}} \subseteq \widetilde{\mathcal{K}}_{\underline{q}}$, we also have $I_{\underline{q}} \leq \widetilde{I}_{\underline{q}}$. Define the maximally mixed state on $\widetilde{\mathcal{K}}_{\underline{q}}$ to be:

$$\tau_{\underline{q}} := \frac{\widetilde{I}_{\underline{q}}}{|\widetilde{\mathcal{K}}_{\underline{q}}|}, \tag{15}$$

where $|\widetilde{\mathcal{K}}_{\underline{q}}|$ is the dimension of the space $\widetilde{\mathcal{K}}_{\underline{q}}$. The following inequality holds:

$$|\widetilde{\mathcal{K}}_{\underline{q}}| \leq (n+1)^{d^2} |\mathcal{K}_{\underline{q}}|. \tag{16}$$

For sake of completeness, we provide a proof of the above inequality in Appendix A. Further define

$$\tau_n := \frac{1}{|\mathcal{P}^n_{\mathcal{Y}}|} \sum_{\underline{q} \in \mathcal{P}^n_{\mathcal{Y}}} \tau_{\underline{q}}. \tag{17}$$

Note that τ_n does not depend on the choice of the initial basis $\{|y^n\rangle\}$. Consider a state σ whose spectral decomposition is given by

$$\sigma = \sum_{y=1}^{d} \lambda_y |y\rangle\langle y|. \tag{18}$$

Then $S(\sigma) = H(\underline{\lambda})$, where $\underline{\lambda} := (\lambda_1, \cdots, \lambda_d)$.

Lemma 5. *For any state $\sigma \in \mathcal{B}(\mathcal{H})$,*

$$(n+1)^{(d^2+d)} \tau_n \geq \sigma^{\otimes n}. \tag{19}$$

Proof. Assume that the spectral decomposition of σ is given by (18). Then[32]

$$\sigma^{\otimes n} = \sum_{\underline{q} \in \mathcal{P}_y^n} 2^{-n[D(\underline{q}\|\underline{\lambda})+H(\underline{q})]} I_{\underline{q}}, \tag{20}$$

where $D(\underline{q}\|\underline{\lambda}) := \sum_{i=1}^{d} q_i \log \frac{q_i}{\lambda_i}$ denotes the relative entropy. Since

$$
\begin{aligned}
I_{\underline{q}} \sigma^{\otimes n} I_{\underline{q}} &= 2^{-n[D(\underline{q}\|\underline{\lambda})+H(\underline{q})]} I_{\underline{q}} \\
&\leq 2^{-nD(\underline{q}\|\underline{\lambda})} \frac{I_{\underline{q}}}{|\mathcal{K}_{\underline{q}}|} \\
&\leq (n+1)^{d^2} \frac{\widetilde{I}_{\underline{q}}}{|\widetilde{\mathcal{K}}_{\underline{q}}|} \\
&= (n+1)^{d^2} \tau_{\underline{q}}.
\end{aligned}
\tag{21}
$$

In the above, the first inequality follows from the following fact (similar to (8)): $|\mathcal{K}_{\underline{q}}| = |\mathcal{T}_y^n(\underline{q})| \leq 2^{nH(\underline{q})}$ for all $\underline{q} \in \mathcal{P}_y^n$. The second inequality follows from the non-negativity of $D(\underline{q}\|\underline{\lambda})$, $I_{\underline{q}} \leq \widetilde{I}_{\underline{q}}$, and (16). The final equality follows from the definition (15) of $\tau_{\underline{q}}$. Then

$$
\begin{aligned}
\sigma^{\otimes n} &= \sum_{\underline{q} \in \mathcal{P}_y^n} I_{\underline{q}} \sigma^{\otimes n} I_{\underline{q}} \\
&\leq (n+1)^{d^2} \sum_{\underline{q} \in \mathcal{P}_y^n} \tau_{\underline{q}} \\
&= (n+1)^{d^2} |\mathcal{P}_y^n| \tau_n \\
&\leq (n+1)^{(d^2+d)} \tau_n,
\end{aligned}
$$

where the first inequality follows from (21), the next identity follows from (17), and the second inequality follows from the fact (similar to (6)): $|\mathcal{P}_y^n| \leq (n+1)^d$. $\qquad\square$

9

Let $x_o^n = (\mathsf{x}_1, \ldots, \mathsf{x}_1, \mathsf{x}_2, \ldots, \mathsf{x}_2, \ldots, \mathsf{x}_k, \ldots, \mathsf{x}_k)$ be the ordered sequence in $\mathcal{T}_{\mathcal{X}}^n(\underline{q})$, where the number of x_i in x_o^n is $N(\mathsf{x}_i | x_o^n) = m_i = nq_i$. Define

$$\omega_{x_o^n} := \tau_{m_1} \otimes \cdots \otimes \tau_{m_k}, \tag{22}$$

where each τ_{m_i} is defined similarly to (17).

For any $x^n \in \mathcal{T}_{\mathcal{X}}^n(\underline{q})$, there exists a permutation $s \in S_n$ such that $x^n = sx_o^n$. Let U_s be the unitary representation of s in $\mathcal{H}^{\otimes n}$. We can then define the following state that plays an important role in the following sections:

$$\omega_{x^n} = U_s \omega_{x_o^n} U_s^\dagger. \tag{23}$$

We define the δ-typical projector $\Pi_{\sigma,\delta}^n$ of $\sigma^{\otimes n}$ to be

$$\Pi_{\sigma,\delta}^n = \sum_{y^n \in \mathcal{T}_{\underline{\lambda},\delta}^n} |y^n\rangle\langle y^n|,$$

where σ is defined in (18) and $\mathcal{T}_{\underline{\lambda},\delta}^n$ is similarly defined as $\mathcal{T}_{\underline{p},\delta}^n$ in (9).

For any $\varepsilon, \delta > 0$, some positive constant $c := H(\underline{p})$ depending only on $\underline{p}$, and sufficiently large n, we have[33]:

$$\operatorname{Tr} \sigma^{\otimes n} \Pi_{\sigma,\delta}^n \geq 1 - \varepsilon \tag{24}$$

$$2^{-n[S(\sigma)+c\delta]} \Pi_{\sigma,\delta}^n \leq \Pi_{\sigma,\delta}^n \sigma^{\otimes n} \Pi_{\sigma,\delta}^n \leq 2^{-n[S(\sigma)-c\delta]} \Pi_{\sigma,\delta}^n \tag{25}$$

$$\operatorname{Tr} \Pi_{\sigma,\delta}^n \leq 2^{n[S(\sigma)+c\delta]}. \tag{26}$$

III. UNIVERSAL PACKING

Consider a classical-quantum channel $W^B : x \to W^B(x)$ which maps the input alphabets $\mathcal{X}$ (with probability distribution $\underline{p}$ on $\mathcal{X}$) to the set of states in $\mathfrak{D}(\mathcal{H}_B)$, (where $\dim \mathcal{H}_B = d_B$). Then using this channel n times gives a memoryless channel $W^{B^n} \equiv (W^B)^{\otimes n}$ that maps $x^n \in \mathcal{X}^n$ to a tensor product state in $\mathfrak{D}(\mathcal{H}_B^{\otimes n})$:

$$W^{B^n}(x^n) := W^B(x_1) \otimes \cdots \otimes W^B(x_n). \tag{27}$$

Suppose Alice wants to send classical messages in the set $\mathcal{M}_n := \{1, 2, \ldots, M_n\}$ to Bob through the channel W^{B^n}. In order to do this she needs to encode her messages appropriately before sending them through the channel and Bob needs a decoder to decode the messages

that he receives. The encoding performed by Alice is a map φ_n from the set of messages $\mathcal{M}_n$ to a set $\mathcal{A}_n \subset \mathcal{X}^n$. The decoding performed by Bob is a POVM $\Upsilon^n := \{\Upsilon_i\}_{i=1}^{M_n}$, where each POVM element Υ_i is an operator acting on $\mathcal{H}_B^{\otimes n}$.

A "c-q" code $\mathcal{C}_n(W^B)$ is given by the triple $\mathcal{C}_n(W^B) := \{M_n, \varphi_n, \Upsilon^n\}$, where M_n denotes the size of the code (i.e., the number of codewords). The *average error probability* of the code $\mathcal{C}_n(W^B)$ is given by:

$$p_e\left(\mathcal{C}_n(W^B)\right) := \frac{1}{M_n} \sum_{i=1}^{M_n} \mathrm{Tr}\, W^{B^n}(\varphi_n(i))\,(\mathbb{I} - \Upsilon_i). \tag{28}$$

Let $\mathcal{C}(W^B) := \{\mathcal{C}_n(W^B)\}_{n=1}^{\infty}$ denote a sequence of such c-q codes. For such a sequence of codes, a real number R,

$$R := \lim_{n \to \infty} \frac{1}{n} \log M_n \tag{29}$$

is called an achievable rate if

$$p_e\left(\mathcal{C}_n(W^B)\right) \to 0 \quad \text{as} \quad n \to \infty.$$

We will refer to a sequence of codes simply as a code when there is no possibility of ambiguity.

It has been shown that[1,2] for every classical-quantum channel W^B, and any probability distribution $\underline{p}$ on $\mathcal{X}$, there exists a sequence of c-q codes $\mathcal{C}(W^B)$ with achievable rate

$$\chi(\underline{p}, W^B) := S(\overline{W}^B) - \sum_{i=1}^{k} p_i S(W^B(\mathsf{x}_i)),$$

where

$$\overline{W}^B = \sum_{i=1}^{k} p_i W^B(\mathsf{x}_i). \tag{30}$$

Note that $\chi(\underline{p}, W^B)$ is just the Holevo quantity of the output ensemble $\{p_x, W^B(x)\}_{x \in \mathcal{X}}$ of the classical-quantum channel. However, the decoding POVM Υ^n of the code $\mathcal{C}_n(W^B)$ constructed requires knowledge of the channel W^{B}[1,2].

Hayashi[14] proved that it is possible to construct a decoding POVM that is independent of a given channel W^B. He then constructed a sequence of universal c-q codes with achievable rate approaching $\chi(\underline{p}, W^B)$ by combining Schur-duality and a classical universal code proposed by Csiszár and Körner. We use a decoding POVM similar to Hayashi's and construct a different sequence of universal c-q codes. Unlike Hayashi's methods of code construction, we employ "type decomposition" and the regular random coding technique, though certain parts of the proof are similar to those used by Hayashi[14].

Theorem 1. *Given a probability distribution $\underline{p}$ on the input alphabet $\mathfrak{X}$, of any classical-quantum channel W^B, there exists a sequence, $\mathfrak{C}(W^B)$, of codes which can achieve any rate $R < \chi(\underline{p}, W^B)$, and are universal, in the sense that their decoding POVMs only depend on $\chi(\underline{p}, W^B)$ and not on specific knowledge of the structure of the channel W^B.*

Proof. We will employ the random coding technique to show the existence of such a universal code $\mathfrak{C}$.

Let $M_n = |\mathcal{M}_n|$. Let $\mathcal{A}_n := \{X_i\}_{i=1}^{M_n}$, where each X_i is a random variable chosen independently, according to

$$p'^n_{x^n} := \Pr\{X_i = x^n\} = \begin{cases} p^n_{x^n}/Q_n, & \text{if, } x^n \in \mathfrak{T}^n_{\underline{p},\delta} \\ 0 & \text{otherwise,} \end{cases} \tag{31}$$

where Q_n is defined in (10). It is easy to verify that

$$\|p - p'\|_1 = \sum_{x^n} |p^n_{x^n} - p'^n_{x^n}| \leq 2\varepsilon. \tag{32}$$

The codeword $\varphi_n(i)$ for the i-th message is given by the realization of the random variable X_i, taking values $x^n \in \mathfrak{T}^n_{\underline{p},\delta}$.

To construct a suitable POVM, we define the projector similar to Hayashi's

$$\Lambda_{X_i} := \{\omega_{X_i} - 2^{n\gamma_n}\tau_n \geq 0\}, \tag{33}$$

where γ_n is a real number to be determined below.

Define the POVM element Υ_i:

$$\Upsilon_i := \left(\sum_{j=1}^{M_n} \Lambda_{X_j}\right)^{-1/2} \Lambda_{X_i} \left(\sum_{j=1}^{M_n} \Lambda_{X_j}\right)^{-1/2}, \tag{34}$$

with Λ_{X_i} being given by (33). Apparently, each Λ_{X_i} (and therefore each POVM element Υ_i) does not depend on full knowledge of the channel W^B. We then define the random universal code $\mathfrak{C}^n_{\text{rc}} := \{X_i, \Upsilon_i\}_{i=1}^{M_n}$.

Using the following operator inequality[34]:

$$\mathbb{I} - \sqrt{S+T}^{-1} S \sqrt{S+T}^{-1} \leq 2(\mathbb{I} - S) + 4T,$$

where $0 \leq S \leq \mathbb{I}$ and $T \geq 0$, the average error probability $p_e(\mathfrak{C}^n_{\text{rc}})$ in (28) can be bounded as follows:

$$p_e(\mathfrak{C}^n_{\text{rc}}) \leq \frac{2}{M_n} \sum_{i=1}^{M_n} \text{Tr}\left(\mathbb{I} - \Lambda_{X_i}\right) W^{B^n}(X_i) + \frac{4}{M_n} \sum_{i=1}^{M_n} \sum_{j=1, j\neq i}^{M_n} \text{Tr}\,\Lambda_{X_j} W^{B^n}(X_i). \tag{35}$$

Taking the expectation of (35), with respect to the distribution (31), we get

$$\mathbb{E}\left[p_e(\mathcal{C}_{\mathrm{rc}}^n)\right] \leq \frac{2}{M_n} \sum_{i=1}^{M_n} \mathbb{E}_i \left[\mathrm{Tr}\left(\mathbb{I} - \Lambda_{X_i}\right) W^{B^n}(X_i) \right] + \frac{4}{M_n} \sum_{i=1}^{M_n} \sum_{\substack{j=1 \\ j \neq i}}^{M_n} \mathbb{E}_{ij} \left[\mathrm{Tr}\, \Lambda_{X_j} W^{B^n}(X_i) \right]. \quad (36)$$

We evaluate the sum in the first term of the above inequality in Sec. III A and the inner sum in the second term in Sec. III B. These results (in particular, (47) and (52) below) yield the following upper bound on $\mathbb{E}\left[p_e(\mathcal{C}_{\mathrm{rc}}^n)\right]$: For any $0 < t < 1$ we have that

$$\mathbb{E}\left[p_e(\mathcal{C}_{\mathrm{rc}}^n)\right] \leq 2^{-nt\left[\chi_{1-t} - \gamma_n - \zeta_n(k(d_B^2 + d_B))\right]} + 4(1-\varepsilon)^{-1} M_n 2^{-n(\gamma_n - \zeta_n(d_B^2 + d_B))} + 2\varepsilon \quad (37)$$

where $\chi_{1-t} \equiv \chi_{1-t}(\{p_x, W^B(x)\})$ is the the $(1-t)$-χ quantity of the ensemble $\mathcal{E} := \{p_x, W^B(x)\}$, given by (3), and $\zeta_n(\cdot)$ is defined through (7). Denote

$$\varepsilon_n := 2^{-nt\left[\chi_{1-t} - \gamma_n - \zeta_n(k(d_B^2 + d_B))\right]} + 4(1-\varepsilon)^{-1} M_n 2^{-n(\gamma_n - \zeta_n(d_B^2 + d_B))} + 2\varepsilon. \quad (38)$$

Our aim is to show that by choosing γ_n appropriately, we can ensure that $\varepsilon_n \to 0$ as $n \to \infty$ for any $R < \chi(\underline{p}, W^B)$. This implies in particular, that if all that is known about the classical-quantum channel is the probability distribution $\underline{p}$ on $\mathcal{X}$ and a lower bound (say χ_0) to the value of the corresponding Holevo quantity, $\chi(\underline{p}, W^B)$, then there exists a universal code $\mathcal{C}(W^B)$, which can achieve any rate $R < \chi_0$.

Let

$$M_n = 2^{n\left[R - \zeta_n((k+1)(d_B^2 + d_B))\right]}. \quad (39)$$

Note that this choice respects the definition (29) of a rate R, since $\zeta_n((k+1)(d_B^2 + d_B)) \to 0$ as $n \to \infty$. Further, for any $t \in (0, 1)$, let us choose

$$\gamma_n = R + r(t) - \zeta_n(k(d_B^2 + d_B)), \quad (40)$$

where we define[14]

$$r(t) := \frac{t}{t+1}(\chi_{1-t} - R). \quad (41)$$

This choice of γ_n reduces the second term on the RHS of (37) to $4(1-\varepsilon)^{-1} \times 2^{-nr(t)}$, and the first term on the RHS of (37) to

$$2^{-nt[\chi_{1-t} - R - r(t) + \zeta_n(k(d_B^2 + d_B)) - \zeta_n(k(d_B^2 + d_B))]}$$

$$= 2^{-nt[\chi_{1-t} - R - (\chi_{1-t} - R - \frac{r(t)}{t})]}$$

$$= 2^{-nr(t)}, \quad (42)$$

13

where we have used the fact that $r(t)$, defined by (41), is equivalently expressed as $r(t) = \chi_{1-t} - R - r(t)/t$. Hence we obtain the following:

$$\mathbb{E}\left[p_e(\mathcal{C}_{\mathrm{rc}}^n)\right] \leq (1 + 4(1 - \varepsilon)^{-1}) \times 2^{-nr(t)} + 2\varepsilon. \tag{43}$$

Since this holds for any $t \in (0, 1)$, we have in particular that

$$\mathbb{E}\left[p_e(\mathcal{C}_{\mathrm{rc}}^n)\right] \leq (1 + 4(1 - \varepsilon)^{-1}) \times 2^{-n \max_{t \in (0,1)} r(t)} + 2\varepsilon. \tag{44}$$

To prove that $\mathbb{E}\left[p_e(\mathcal{C}_{\mathrm{rc}}^n)\right]$ vanishes asymptotically for a suitable range of values of R, it suffices to prove that $\max_{t \in (0,1)} r(t) > 0$ for that range of values of R. From (4) we infer that

$$\lim_{t \searrow 0} \chi_{1-t} = \chi, \tag{45}$$

where $\chi = \chi(\underline{p}, W^B)$. Moreover, the convergence is from below, since $t \mapsto \chi_{1-t}$ is monotonically decreasing. This ensures that for any $R < \chi$, there exists a $t_R > 0$ such that

$$\chi_{1-t} - R > 0, \quad \forall\, t < t_R.$$

This in turn implies that for any given $R < \chi$,

$$\max_{t \in (0,1)} r(t) > 0. \tag{46}$$

$\square$

Thus there must exist a sequence of codes $\mathcal{C}_n$ of rate $R < \chi$ such that $p_e(\mathcal{C}_n) \to 0$ as $n \to \infty$. It follows that for any $\varepsilon > 0$, and sufficiently large n,

$$p_e(\mathcal{C}_n) = \frac{1}{M_n} \sum_{i=1}^{M_n} p_e(i) < \varepsilon,$$

where M_n is given by (39) and $p_e(i) := \mathrm{Tr}\, W^{B^n}(\varphi_n(i))(\mathbb{I} - \Upsilon_i)$ denotes the probability of error corresponding to the ith message.

A. Evaluation of the first term in (36)

In the section, we closely follow[14].

Lemma 6.

$$\mathbb{E}_i\Big[\mathrm{Tr}(\mathbb{I} - \Lambda_{X_i})W^{B^n}(X_i)\Big] \leq 2^{-nt\big[\chi_{1-t}-\gamma_n-\zeta_n(k(d_B^2+d_B))\big]} + 2\varepsilon \tag{47}$$

where $\chi_{1-t} \equiv \chi_{1-t}(\{p_x, W^B(x)\})$ *is the* $(1-t)$-χ *quantity of the ensemble* $\mathcal{E} := \{p_x, W^B(x)\}$, *defined through (3), and* $\zeta_n(k(d_B^2 + d_B))$ *is defined through (7).*

Proof. Consider the projector Λ_{x^n} defined in (33), where $\gamma_n \in \mathbb{R}$. Since $[\omega_{x^n}, \tau_n] = 0$, and this in turn can be shown to imply that for any $t \in (0,1)^{14}$:

$$(\mathbb{I} - \Lambda_{x^n}) \leq \omega_{x^n}^{-t} 2^{nt\gamma_n} \tau_n^t. \tag{48}$$

For sake of completeness, the commutativity of the operators ω_{x^n} and τ_n is proved in Appendix B. Since $W^{B^n}(x_o^n) = (W^B(\mathsf{x}_1))^{\otimes m_1} \otimes \cdots \otimes (W^B(\mathsf{x}_k))^{\otimes m_k}$, by direct application of Lemma 5, we obtain

$$
\begin{aligned}
W^{B^n}(x^n) &= U_s W^{B^n}(x_o^n) U_s^\dagger \\
&\leq (n+1)^{k(d_B^2+d_B)} U_s\left(\tau_{m_1} \otimes \ldots \otimes \tau_{m_k}\right) U_s^\dagger \\
&= (n+1)^{k(d_B^2+d_B)} U_s \omega_{x_o^n} U_s^\dagger \\
&= (n+1)^{k(d_B^2+d_B)} \omega_{x^n}.
\end{aligned}
\tag{49}
$$

This yields, for any $t \in (0,1)$:

$$
\begin{aligned}
\omega_{x^n}^{-t} &\leq (n+1)^{tk(d_B^2+d_B)} \left(W^{B^n}(x^n)\right)^{-t} \\
&= 2^{nt\zeta_n(k(d_B^2+d_B))} \left(W^{B^n}(x^n)\right)^{-t},
\end{aligned}
\tag{50}
$$

and

$$W^{B^n}(x^n)\omega_{x^n}^{-t} \leq 2^{nt\zeta_n(k(d_B^2+d_B))} \left(W^{B^n}(x^n)\right)^{1-t}. \tag{51}$$

Finally,

$$
\begin{aligned}
\mathbb{E}_i\Big[\mathrm{Tr}(\mathbb{I} - \Lambda_{X_i})W^{B^n}(X_i)\Big] &= \sum_{x^n \in \mathcal{T}_{\underline{p},\delta}^n} p_{x^n}'^n \left[\mathrm{Tr}(\mathbb{I} - \Lambda_{x^n})W^{B^n}(x^n)\right] \\
&\leq \sum_{x^n \in \mathcal{X}^n} p_{x^n}^n \left[\mathrm{Tr}(\mathbb{I} - \Lambda_{x^n})W^{B^n}(x^n)\right] + 2\varepsilon \\
&\leq 2^{nt\gamma_n}\, \mathrm{Tr}\left(\sum_{x^n \in \mathcal{X}^n} p_{x^n}^n \omega_{x^n}^{-t} W^{B^n}(x^n)\right)\tau_n^t + 2\varepsilon \\
&\leq 2^{nt\zeta_n(k(d_B^2+d_B))} 2^{nt\gamma_n}\left[\mathrm{Tr}\left(\sum_{x^n \in \mathcal{X}^n} p_{x^n}^n \left(W^{B^n}(x^n)\right)^{1-t}\right)\tau_n^t\right] + 2\varepsilon \\
&\leq 2^{nt\zeta_n(k(d_B^2+d_B))} 2^{nt\gamma_n} \max_\sigma\left[\mathrm{Tr}\Big(\sum_{x \in \mathcal{X}} p_x \left(W^B(x)\right)^{1-t}\Big)^{\otimes n}\sigma^t\right] + 2\varepsilon.
\end{aligned}
$$

15

The first equality follows from evaluating the expectation. The first inequality follows from (32). The second inequality follows from (48). The third inequality follows from (51).

Applying Lemma 4 to the above equation, we get

$$
= 2^{nt[\zeta_n(k(d_B^2+d_B))+\gamma_n]} \left\{ \mathrm{Tr}\left[\left(\sum_{x \in \mathcal{X}} p_x \left(W^B(x) \right)^{1-t} \right)^{\otimes n} \right]^{\frac{1}{1-t}} \right\}^{1-t} + 2\varepsilon
$$

$$
= 2^{nt[\zeta_n(k(d_B^2+d_B))+\gamma_n]} \left\{ \mathrm{Tr}\left[\sum_{x \in \mathcal{X}} p_x \left(W^B(x) \right)^{1-t} \right]^{\frac{1}{1-t}} \right\}^{n(1-t)} + 2\varepsilon
$$

$$
= 2^{-nt\left[\chi_{1-t} - \gamma_n - \zeta_n(k(d_B^2+d_B)) \right]} + 2\varepsilon
$$

where $\chi_{1-t} \equiv \chi_{1-t}(\{p_x, W^B(x)\})$. □

B. Evaluation of the second term in (36)

We have, for given $i, j \in \{1, 2, \ldots, M_n\}$ and $i \neq j$,

$$
\mathbb{E}_{i,j}\left[\mathrm{Tr}\,\Lambda_{X_j} W^{B^n}(X_i) \right] = \mathbb{E}_j\left[\mathrm{Tr}\,\Lambda_{X_j}\mathbb{E}_i\left[W^{B^n}(X_i) \right] \right]
$$

$$
= \mathbb{E}_j\left[\mathrm{Tr}\left(\Lambda_{X_j}\frac{1}{Q_n} \sum_{x^n \in \mathcal{T}_{p,\delta}^n} p_{x^n}^n W^{B^n}(x^n) \right) \right]
$$

$$
\leq (1-\varepsilon)^{-1}\mathbb{E}_j\,\mathrm{Tr}\left(\Lambda_{X_j}\left(\overline{W}^B \right)^{\otimes n} \right)
$$

$$
\leq (1-\varepsilon)^{-1}(n+1)^{(d_B^2+d_B)}\mathbb{E}_j\,\mathrm{Tr}\left[\Lambda_{X_j}\tau_n \right]
$$

$$
\leq (1-\varepsilon)^{-1}2^{n\zeta_n(d_B^2+d_B)}2^{-n\gamma_n}
$$

$$
= (1-\varepsilon)^{-1}2^{-n\left[\gamma_n - \zeta_n(d_B^2+d_B) \right]}. \tag{52}
$$

In the above, the first inequality follows from that fact $Q_n \geq 1 - \varepsilon$ and

$$
\left(\overline{W}^B \right)^{\otimes n} = \sum_{x^n \in \mathcal{X}^n} p_{x^n}^n W^{B^n}(x^n)
$$

$$
\geq \sum_{x^n \in \mathcal{T}_{p,\delta}^n} p_{x^n}^n W^{B^n}(x^n). \tag{53}
$$

The second inequality follows from Lemma 5. The third inequality follows from Lemma 1. $\zeta_n(d_B^2 + d_B)$ in the last equality is defined in (7).

16

IV. UNIVERSAL COVERING

Consider the probability distribution $\underline{p} = (p_1, \cdots, p_k)$ on $\mathfrak{X}$ and a classical-quantum channel $W^E : x \in \mathfrak{X} \to W^E(x) \in \mathfrak{D}(\mathcal{H}_E)$, with $d_E = \dim \mathcal{H}_E$. Then using this channel n times gives a memoryless channel $(W^E)^{\otimes n} := W^{E^n}$ that maps a sequence $x^n \in \mathfrak{X}^n$ with probability $p_{x^n}^n := p_{x_1} \cdots p_{x_n}$ to a product state in $\mathfrak{D}(\mathcal{H}_E^{\otimes n})$:

$$W^{E^n}(x^n) = W^E(x_1) \otimes \cdots \otimes W^E(x_n).$$

Let $\overline{W}^E = \sum_{i=1}^k p_i W^E(\mathsf{x}_i)$. Then

$$\overline{W}^{E^n} := (\overline{W}^E)^{\otimes n} = \sum_{x^n \in \mathfrak{X}^n} p_{x^n}^n W^{E^n}(x^n).$$

Consider a subset $\mathcal{S} \subset \mathfrak{X}^n$, and define the "obfuscation error"

$$\Delta(\mathcal{S}) := \left\| \frac{1}{|\mathcal{S}|} \sum_{x^n \in \mathcal{S}} W^{E^n}(x^n) - \overline{W}^{E^n} \right\|_1. \tag{54}$$

We are interested in finding the smallest "covering" subset $\mathcal{S} \subset \mathfrak{X}^n$ for which $\Delta(\mathcal{S}) \to 0$ as $n \to \infty$. We discover that for any given probability distribution $\underline{p}$, the covering set is universal in the sense that its size $|\mathcal{S}|$ depends only on the value of $\chi(\underline{p}, W^E)$, and not on any specific knowledge of the channel W^E itself. Here we provide an explicit proof of the fact that $\Delta(\mathcal{S})$ can be made arbitrarily small for any randomly picked subset $\mathcal{S} \subset \mathfrak{X}^n$ as long as n is sufficiently large and $\frac{1}{n} \log |\mathcal{S}| > \chi(\underline{p}, W^E)$[18], a result which follows from the so-called "covering lemma"[17,35,36]. More precisely, given any upper bound (say χ_1) on the Holevo quantity $\chi(\underline{p}, W^E)$, there is a subset $\mathcal{S}$ of the δ-typical set, $\mathfrak{T}_{\underline{p},\delta}^n$ (defined through (9)) for any $\delta > 0$, for which $\Delta(\mathcal{S}) \to 0$ as $n \to \infty$ provided $|\mathcal{S}| > \chi_1$.

Theorem 2. *For any $\varepsilon, \delta > 0$, let $L_n = 2^{n[\chi_1 + 2c\delta]}$, where χ_1 is a given upper bound on the Holevo quantity $\chi(\underline{p}, W^E)$. Then the set $\mathcal{A}_n = \{X_i\}_{i=1}^{L_n}$, where each X_i is a random variable chosen independently according to*

$$p_{x^n}^{\prime n} := \Pr\{X_i = x^n\} = \begin{cases} p_{x^n}^n / Q_n, & \text{if } x^n \in \mathfrak{T}_{\underline{p},\delta}^n, \\ 0, & \text{otherwise} \end{cases} \tag{55}$$

(Q_n is given in (10)), satisfies

$$\Pr\left\{ \Delta(\mathcal{A}_n) \geq \varepsilon + 4\sqrt{k\varepsilon} + 8\sqrt{3\varepsilon + 2\sqrt{k\varepsilon}} \right\} \leq \varepsilon_n, \tag{56}$$

for a positive constant ε_n (given in (73)) such that $\varepsilon_n \to 0$ as $n \to \infty$.

Proof. Recall that for any sequence $x^n \in \mathcal{X}^n$, there exists a permutation $s \in S_n$ such that $x^n = sx_o^n$, where x_o^n denotes the ordered sequence corresponding to x^n. Let U_s be the unitary representation of s in $\mathcal{H}_E^{\otimes n}$, and let us define

$$\Pi_{W^E,\delta}^n(x^n) = U_s \Pi_{W^E,\delta}^n(x_o^n) U_s^\dagger,$$

where $\Pi_{W^E,\delta}^n(x_o^n)$ is the δ-conditional typical projector for the state $W^{E^n}(x_o^n)$:

$$\Pi_{W^E,\delta}^n(x_o^n) \equiv \Pi_{W^E(\mathsf{x}_1),\delta}^{m_1} \otimes \cdots \otimes \Pi_{W^E(\mathsf{x}_k),\delta}^{m_k},$$

and $\Pi_{W^E(\mathsf{x}_i),\delta}^{m_i}$ is the δ-typical projector for the state $W^{E^{m_i}}(\mathsf{x}_i)$ such that

$$\operatorname{Tr} \Pi_{W^E(\mathsf{x}_i),\delta}^{m_i} W^{E^{m_i}}(\mathsf{x}_i) \geq 1 - \varepsilon.$$

Let

$$\sigma(x^n) := \Pi_{W^E,\delta}^n(x^n) W^{E^n}(x^n) \Pi_{W^E,\delta}^n(x^n).$$

Then

$$\begin{aligned}
\operatorname{Tr} \sigma(x^n) &= \operatorname{Tr} \Pi_{W^E,\delta}^n(x^n) W^{E^n}(x^n) \\
&= \operatorname{Tr} \Pi_{W^E,\delta}^n(x_o^n) W^{E^n}(x_o^n) \\
&= \prod_{i=1}^k \operatorname{Tr} \Pi_{W^E(\mathsf{x}_i),\delta}^{m_i} W^{E^{m_i}}(\mathsf{x}_i) \\
&\geq 1 - k\varepsilon.
\end{aligned} \tag{57}$$

Applying the gentle measurement lemma, Lemma 2, to (57) gives

$$\|\sigma(x^n) - W^{E^n}(x^n)\|_1 \leq 2\sqrt{k\varepsilon}. \tag{58}$$

Define the δ-typical projector $\Pi_{\overline{W}^E,\delta}^n$ for the average state $\overline{W}^{E^n}$ such that

$$\operatorname{Tr} \Pi_{\overline{W}^E,\delta}^n \overline{W}^{E^n} \geq 1 - \varepsilon. \tag{59}$$

Define

$$\phi(x^n) = \Pi_{\overline{W}^E,\delta}^n \sigma(x^n) \Pi_{\overline{W}^E,\delta}^n.$$

Then

$$\begin{aligned}
\operatorname{Tr} \phi(x^n) &= \operatorname{Tr} \sigma(x^n) \Pi_{\overline{W}^E,\delta}^n \\
&\geq \operatorname{Tr} \Pi_{\overline{W}^E,\delta}^n W^{E^n}(x^n) - \|\sigma(x^n) - W^{E^n}(x^n)\|_1 \\
&\geq 1 - \varepsilon - 2\sqrt{k\varepsilon},
\end{aligned} \tag{60}$$

18

where the first inequality follows from the fact that for any operator ω, $\|\omega\|_1 = \max_{-\mathbb{I} \leq \Pi \leq \mathbb{I}} \operatorname{Tr} \Pi \omega$, and the second inequality follows from (58) and (59). Applying the gentle measurement lemma to (60) gives

$$\|\phi(x^n) - \sigma(x^n)\|_1 \leq 2\sqrt{\varepsilon + 2\sqrt{k\varepsilon}}. \tag{61}$$

Using the triangle inequality, and the bounds (58) and (61), yields

$$\|\phi(x^n) - W^{E^n}(x^n)\|_1 \leq 2\sqrt{k\varepsilon} + 2\sqrt{\varepsilon + 2\sqrt{k\varepsilon}}. \tag{62}$$

Let

$$\overline{\phi} := \mathbb{E}\phi(X_i) = \sum_{x^n \in \mathcal{T}_{\underline{p},\delta}^n} p_{x^n}^{\prime n} \phi(x^n).$$

Then (60) implies that

$$\operatorname{Tr} \overline{\phi} \geq 1 - \varepsilon - 2\sqrt{k\varepsilon}. \tag{63}$$

Define

$$\Pi := \left\{ \overline{\phi} - \varepsilon 2^{-n[S(\overline{W}^E) + c\delta]} \Pi_{\overline{W}^E, \delta}^n \geq 0 \right\}. \tag{64}$$

Note that $[\overline{\phi}, \Pi_{\overline{W}^E, \delta}^n] = 0$.

Let $\overline{\phi}' := \Pi \overline{\phi} \Pi$. Then

$$\begin{aligned} \operatorname{Tr} \overline{\phi}' &\geq \operatorname{Tr} \overline{\phi} - \varepsilon \\ &\geq 1 - 2\varepsilon - 2\sqrt{k\varepsilon}, \end{aligned} \tag{65}$$

where the first inequality follows from the fact that the support of $\Pi_{\overline{W}^E, \delta}^n$ has dimension less than $2^{n[S(\overline{W}^E) + c\delta]}$, so the eigenvalues smaller than $\varepsilon 2^{-n[S(\overline{W}^E) + c\delta]}$ contribute at most ε to $\operatorname{Tr} \overline{\phi}$, while the second inequality follows from (63). The following operator inequality holds:

$$\overline{\phi}' \geq \varepsilon 2^{-n[S(\overline{W}^E) + c\delta]} \Pi_{\overline{W}^E, \delta}^n \tag{66}$$

where the inequality follows from the definition (64).

For any $x^n \in \mathcal{T}_{\underline{p},\delta}^n$, let $\theta(x^n) := \Pi \phi(x^n) \Pi$. Note that these operators lie in the subspace of the Hilbert space $\mathcal{H}_E^{\otimes n}$ onto which the operator $\Pi_{\overline{W}^E, \delta}^n$ project. Consider the operator ensemble $\{p_{x^n}^{\prime n}, \psi(x^n)\}_{x^n \in \mathcal{T}_{\underline{p},\delta}^n}$, where

$$\psi(x^n) := 2^{n\left[\sum_{i=1}^k p_i S(W^E(\mathsf{x}_i)) - c\delta\right]} \theta(x^n). \tag{67}$$

19

Note that

$$\mathbb{E}\psi(X_i) = 2^{n\left[\sum_{i=1}^{k} p_i S(W^E(\mathsf{x}_i)) - c\delta\right]}\mathbb{E}\theta(X_i), \tag{68}$$

and

$$\mathbb{E}\theta(X_i) = \Pi\overline{\phi}\Pi = \overline{\phi}'. \tag{69}$$

By (68), (69) and (66), we have

$$\mathbb{E}\psi(X_i) \geq \varepsilon 2^{-n\left[\chi(\underline{p},W^E) + 2c\delta\right]}\Pi^n_{\overline{W}^E,\delta}. \tag{70}$$

Hence, the operator ensemble $\{p'^n_{x^n}, \psi(x^n)\}_{x^n \in \mathcal{T}^n_{\underline{p},\delta}}$ satisfies the condition of Lemma 3 with

$$t = \varepsilon 2^{-n\left[\chi(\underline{p},W^E) + 2c\delta\right]}. \tag{71}$$

Applying Lemma 3, we have

$$\Pr\left\{\frac{1}{L_n}\sum_{i=1}^{L_n}\theta(X_i) \notin [1 \pm \varepsilon]\overline{\phi}'\right\} \leq 2\operatorname{Tr}(\Pi^n_{\overline{W}^E,\delta})2^{-L_n k\varepsilon^3 2^{-n[\chi(\underline{p},W^E) + 2c\delta]}} := \varepsilon'_n \tag{72}$$

where $k = 1/(2(\ln 2)^2)$, $L_n = 2^{n[\chi_1 + 2c\delta]}$ and

$$\varepsilon'_n := 2 \times 2^{-k\varepsilon^3 2^{n[\chi_1 - \chi(\underline{p},W^E)]} + n[S(\overline{W}^E) + c\delta]} \to 0 \tag{73}$$

as $n \to \infty$.

To prove Theorem 2, the above inequality (72) needs to be translated into a statement about the operators $W^{E^n}(x^n)$. To do this, assume that for some set $\mathcal{A}_n \subset \mathcal{T}^n_{\underline{p},\delta}$ such that for any $\varepsilon > 0$ and for n large enough,

$$\frac{1}{|\mathcal{A}_n|}\sum_{x^n \in \mathcal{A}_n}\theta(x^n) \in [1 \pm \varepsilon]\overline{\phi}'.$$

Equivalently,

$$\left\|\frac{1}{|\mathcal{A}_n|}\sum_{x^n \in \mathcal{A}_n}\theta(x^n) - \overline{\phi}'\right\|_1 \leq \varepsilon. \tag{74}$$

We shall show that

$$\Delta(\mathcal{A}_n) \leq \varepsilon + 4\sqrt{k\varepsilon} + 8\sqrt{3\varepsilon + 2\sqrt{k\varepsilon}}. \tag{75}$$

From (74) and (65), we obtain that

$$\operatorname{Tr}\left(\frac{1}{|\mathcal{A}_n|}\sum_{x^n \in \mathcal{A}_n}\theta(x^n)\right) \geq 1 - 3\varepsilon - 2\sqrt{k\varepsilon}. \tag{76}$$

Since $\theta(x^n) = \Pi\phi(x^n)\Pi$, applying the gentle measurement lemma to (76) gives

$$\left\|\frac{1}{|\mathcal{A}_n|} \sum_{x^n \in \mathcal{A}_n} \phi(x^n) - \frac{1}{|\mathcal{A}_n|} \sum_{x^n \in \mathcal{A}_n} \theta(x^n)\right\|_1 \leq 2\sqrt{3\varepsilon + 2\sqrt{k\varepsilon}}. \tag{77}$$

Likewise, applying the gentle measurement lemma to (65) gives

$$\|\overline{\phi}' - \overline{\phi}\|_1 \leq 2\sqrt{2\varepsilon + 2\sqrt{k\varepsilon}}. \tag{78}$$

We also have

$$\left\|\frac{1}{|\mathcal{A}_n|} \sum_{x^n \in \mathcal{A}_n} \phi(x^n) - \frac{1}{|\mathcal{A}_n|} \sum_{x^n \in \mathcal{A}_n} W^{E^n}(x^n)\right\|_1$$

$$\leq \frac{1}{|\mathcal{A}_n|} \sum_{x^n \in \mathcal{A}_n} \left\|\phi(x^n) - W^{E^n}(x^n)\right\|_1$$

$$\leq 2\sqrt{k\varepsilon} + 2\sqrt{\varepsilon + 2\sqrt{k\varepsilon}}, \tag{79}$$

where the second inequality follows from (62). We can analogously obtain

$$\|\overline{\phi} - \overline{W}^{E^n}\|_1 \leq 2\sqrt{k\varepsilon} + 2\sqrt{\varepsilon + 2\sqrt{k\varepsilon}}. \tag{80}$$

By applying the triangle inequality and combining (79), (77), (74), (78) and (80) we obtain the desired bound (75) for $\Delta(\mathcal{A}_n)$. The statement of the theorem follows immediately from (72).

$\square$

V.   UNIVERSAL PRIVATE CODING

Consider a $c \to qq$ channel $W^{BE} : x \to W^{BE}(x)$ that maps the input alphabet $\mathcal{X}$ (with the probability distribution $\underline{p}$ on $\mathcal{X}$) to the set of densities in $\mathfrak{D}(\mathcal{H}_B \otimes \mathcal{H}_E)$. Such a channel induces a classical-quantum channel $W^B : x \to W^B(x)$, where $W^B(x) = \mathrm{Tr}_E W^{BE}(x)$, from the sender Alice to the receiver Bob. Meanwhile, it also induces a classical-quantum channel $W^E : x \to W^E(x)$, where $W^E(x) = \mathrm{Tr}_B W^{BE}(x)$, from the sender Alice to an eavesdropper Eve.

The communication task is for Alice to send a private classical message in the set $\mathcal{J}_n := \{1, 2, \cdots, J_n\}$ reliably to the receiver Bob through n uses of the $c \to qq$ channel, i.e., through $W^{B^n E^n} := (W^{BE})^{\otimes n}$, such that Eve cannot obtain any information about the message sent

by Alice. In order for them to achieve this goal, Alice and Bob require an encoder and decoder, respectively. The encoding performed by Alice is a map φ_n that maps a classical message $i \in \mathcal{J}_n$ to an arbitrary element $x^n \in \mathcal{S}_i \subset \mathcal{X}^n$, where each "covering set" $\mathcal{S}_i$ is disjoint and has the same size. The decoding performed by Bob is a POVM $\Upsilon^n := \{\Upsilon_i\}_{i \in \mathcal{J}_n}$, where each POVM element is an operator acting on $\mathcal{H}_B^{\otimes n}$. For any arbitrary $\varepsilon > 0$, we can formally define an (n, ε) "private" code $\mathcal{C}_n(W^{BE}) := \{J_n, \varphi_n, \Upsilon^n\}$ for the channel W^{BE} by

1. Alice's encoding $\varphi_n : i \to x^n \in \mathcal{S}_i \subset \mathcal{X}^n$;

2. Bob's decoding POVM $\Upsilon^n : \mathfrak{D}(\mathcal{H}_B^{\otimes n}) \to \mathcal{J}_n$,

such that for n large enough, the following conditions hold:

- the average error probability of $\mathcal{C}_n(W^{BE})$:

$$p_e(\mathcal{C}_n(W^{BE})) := \frac{1}{J_n} \sum_{i=1}^{J_n} \mathrm{Tr}\left[W^{B^n}(\varphi_n(i))(\mathbb{I} - \Upsilon_i)\right] \leq \varepsilon. \tag{81}$$

- Eve cannot obtain any information on the classical message i by measuring the state $W^{E^n}(x^n) := \mathrm{Tr}_{B^n} W^{B^n E^n}(x^n)$ that she has access to, i.e., $\forall i \in \{1, 2, \ldots, J_n\}$:

$$\Delta(\mathcal{S}_i) := \left\|\frac{1}{|\mathcal{S}_i|} \sum_{x^n \in \mathcal{S}_i} W^{E^n}(x^n) - \overline{W}^{E^n}\right\|_1 \leq \varepsilon, \tag{82}$$

where $\overline{W}^{E^n} := (\overline{W}^E)^{\otimes n}$ and $\overline{W}^E = \sum_{x \in \mathcal{X}} p_x W^E(x)$.

Let $\mathcal{C}(W^{BE}) := \{\mathcal{C}_n(W^{BE})\}_{n=1}^{\infty}$ denote a sequence of such private codes. For such a sequence of codes, a real number R,

$$R := \lim_{n \to \infty} \frac{1}{n} \log J_n \tag{83}$$

is called an achievable rate if

$$p_e\left(\mathcal{C}_n(W^{BE})\right) \to 0 \quad \text{as } n \to \infty \tag{84}$$

$$\forall i, \quad \Delta(\mathcal{S}_i) \to 0 \quad \text{as } n \to \infty. \tag{85}$$

We will refer to a sequence of codes simply as a code when there is no possibility of ambiguity.

It has been shown that[4] for any classical-quantum channel W^{BE}, and any probability distribution $\underline{p}$ on $\mathcal{X}$, there exists a private code $\mathcal{C}(W^{BE})$ (more precisely, a sequence $\mathcal{C}(W^{BE}) := \{\mathcal{C}_n(W^{BE})\}_{n=1}^{\infty}$ of private codes) with achievable rate

$$I_c(\underline{p}, W^{BE}) := \chi(\underline{p}, W^B) - \chi(\underline{p}, W^E). \tag{86}$$

The private code $\mathcal{C}(W^{BE})$ constructed by Devetak[4] requires knowledge of the channel W^{BE}.

Our main result in this paper is to show that one can construct a private code even without the full knowledge of the $c \to qq$ channel W^{BE}. The only prior knowledge required for the code construction is that of the probability distribution on the input alphabet $\mathfrak{X}$, and bounds on the corresponding Holevo quantities $\chi(\underline{p}, W^B)$ and $\chi(\underline{p}, W^E)$.

Theorem 3. *Let W^{BE} denote a classical-quantum channel with input alphabet $\mathfrak{X}$. Given a probability distribution $\underline{p}$ on $\mathfrak{X}$ and positive numbers χ_0 and χ_1, such that $\chi_0 \leq \chi(\underline{p}, W^B)$ and $\chi_1 \geq \chi(\underline{p}, W^E)$, there exists a universal private code $\mathcal{C}(W^{BE})$ which can achieve any rate $R \leq \chi_0 - \chi_1$.*

Proof. The idea of the proof is to combine the universal packing lemma and the universal covering lemma of the previous two sections.

Assume that $\chi_0 > \chi_1$ and define $I_c := \chi_0 - \chi_1$. Note that $I_c \leq I_c(\underline{p}, W^{BE})$, where $I_c(\underline{p}, W^{BE})$ is given by (86). Let $\{M_n\}_{n=1}^\infty$ and $\{L_n\}_{n=1}^\infty$ denote sequences of positive integers such that

$$\lim_{n\to\infty} \frac{1}{n} \log M_n = \chi_0, \tag{87}$$

$$\lim_{n\to\infty} \frac{1}{n} \log L_n = \chi_1 + 2c\delta. \tag{88}$$

In the course of the proof it will become evident that M_n is the size of a c-q code and L_n the size of a covering set.

Let $J_n = \lfloor M_n/L_n \rfloor$, and define the sets $\mathcal{J}_n := \{1, 2, \ldots, J_n\}$ and $\mathcal{L}_n := \{1, 2, \ldots, L_n\}$. Fix $\delta, \varepsilon > 0$. We first construct a random code $\mathcal{C}_{\mathrm{rc}}^n$ whose codewords are given by realizations of the random variables in the set $\mathcal{A}_n := \{X_{j,\ell}\}_{j\in[\mathcal{J}_n], \ell\in[\mathcal{L}_n]}$, each $X_{j,\ell}$ being a random variable chosen independently from the δ-typical set, $\mathcal{T}_{\underline{p},\delta}^n$, according to

$$p_{x^n}'^n := \Pr\{X_{j,\ell} = x^n\} = \begin{cases} p_{x^n}^n/Q_n, & \text{if } x^n \in \mathcal{T}_{\underline{p},\delta}^n, \\ 0, & \text{otherwise}, \end{cases} \tag{89}$$

where Q_n is defined in (10).

Note that there are at most M_n pairs of classical indices (j, ℓ). Then by using (87) and the fact that $\chi_0 \leq \chi(\underline{p}, W^B)$, and by invoking Theorem 1, we know that there exists a POVM $\{\Upsilon_{j,\ell}\}$ defined by (34), which would allow Bob to identify the pair of classical indices (j, ℓ) with average error probability $\mathbb{E}\left[p_e(\mathcal{C}_{\mathrm{rc}}^n)\right] < \varepsilon_n$, where ε_n is given by (37) and vanishes

asymptotically with n. The definition of the POVM elements only depends on the value of χ_0 and not on specific knowledge of the structure of the channel W^B.

Define the event

$$\mathbf{I}_0 \equiv \mathbf{I}_0^{(n)} := \left\{ p_e(\mathcal{C}_{\mathrm{rc}}^n) \leq \sqrt{\varepsilon_n} \right\},$$

and let $\mathbf{I}_0^c$ denote the complement of the event $\mathbf{I}_0$. Then using the upper bound on $\mathbb{E}\left[p_e(\mathcal{C}_{\mathrm{rc}}^n)\right]$, and the Markov inequality (see e.g.[37]) for the random variable $p_e(\mathcal{C}_{\mathrm{rc}}^n)$, we have that

$$\begin{aligned}
\Pr\{\mathbf{I}_0^c\} = \Pr\{p_e(\mathcal{C}_{\mathrm{rc}}^n) &> \sqrt{\varepsilon_n}\} \\
&\leq \frac{\mathbb{E}\left[p_e(\mathcal{C}_{\mathrm{rc}}^n)\right]}{\sqrt{\varepsilon_n}} \\
&\leq \sqrt{\varepsilon_n}.
\end{aligned} \tag{90}$$

Let $\mathcal{A}_j^{(n)} = \{X_{j,\ell}\}_{\ell \in \mathcal{L}_n}$, and define the following events:

$$\mathbf{I}_j := \left\{ \Delta(\mathcal{A}_j^{(n)}) < \varepsilon + 4\sqrt{k\varepsilon} + 8\sqrt{3\varepsilon + 2\sqrt{k\varepsilon}} \right\},$$

where $\Delta(\mathcal{A}_j^{(n)})$ denotes the "obfuscation error" of the set $\mathcal{A}_j^{(n)}$, defined as in (54). Then by invoking Theorem 2, we have for each $j = 1, \ldots, J_n$

$$\Pr\{\mathbf{I}_j^c\} \leq \varepsilon_n', \tag{91}$$

for a positive constant ε_n' defined by (73), which vanishes asymptotically with n.

Combining (90) and (91) gives

$$\begin{aligned}
\Pr\{(\mathbf{I}_0 \cap \mathbf{I}_1 \cap \cdots \cap \mathbf{I}_{J_n})^c\} = \Pr\{\mathbf{I}_0^c \cup \mathbf{I}_1^c \cup \cdots \cup \mathbf{I}_{J_n}^c\} \\
\leq \sum_{j=0}^{J_n} \Pr\{\mathbf{I}_j^c\} \leq J_n \varepsilon_n' + \sqrt{\varepsilon_n},
\end{aligned} \tag{92}$$

which goes to 0 as $n \to \infty$. Hence, there exists at least one realization of the set $\mathcal{A}_n$ (and hence of the random code $\mathcal{C}_{\mathrm{rc}}^n$), for which each of the events $\mathbf{I}_j$, for $j = 0, 1, \ldots, J_n$, occurs. Equivalently, (81) and (82) hold.

Note that (92) implies that for n large enough, the event $\mathbf{I}_0 \cap \mathbf{I}_1 \cap \cdots \cap \mathbf{I}_{J_n}$ occurs with high probability, and this in turn ensures that the realizations of the sets $\mathcal{A}_j^{(n)}$ for $j \in \mathcal{J}_n$ are mutually disjoint. Hence, in order to reliably transmit private classical messages to Bob, Alice maps each of her private messages onto a randomly picked element in each different

disjoint set, there being $|\mathcal{J}_n| = J_n$ such sets. She can thus achieve a rate

$$R = \lim_{n \to \infty} \frac{1}{n} \log J_n \le I_c - 2c\delta,$$

Since δ is arbitrary, we arrive at the statement $R \to I_c = \chi_0 - \chi_1$ from below, and there exists at least one particular realization $\mathcal{C}_n = \{x_i^n, \Upsilon_i\}_{i \in \mathcal{J}_n}$ of $\mathcal{C}_{rc}^n$ of rate R such that

$$p_e(\mathcal{C}_n) \le \sqrt{\varepsilon_n}.$$

$\square$

VI. CONCLUSION

We have shown that there exists a universal private coding scheme for a $c \to qq$ channel, W^{BE}, which Alice can use to transmit private messages to Bob (who has access to the system B), at the same time ensuring that an eavesdropper, Eve (who has access to the system E) does not get any information about her messages. The coding scheme only requires knowledge of the probability distribution $\underline{p}$ on the input alphabet, $\mathcal{X}$, of the channel, and of the corresponding bounds on the Holevo quantities of the $c \to q$ channels W^B and W^E, induced between Alice and Bob, and Alice and Eve, respectively. More precisely, a lower bound (say χ_0) on $\chi(\underline{p}, W^B)$, and an upper bound (say χ_1) on $\chi(\underline{p}, W^E)$ suffices. This is because our universal private coding scheme is obtained by combining the universal packing lemma (of Section III) and the universal covering lemma (of Section IV), which depend on these bounds.

Prior knowledge of the probability distribution $\underline{p}$ on the channel's input alphabet seems crucial both in Hayashi's coding scheme[14], as well as in ours. In contrast, the classical results on universal packing and universal covering[15] do not require this knowledge. For the universal private coding, we also require knowledge of the values of χ_0 and χ_1 individually. It would be interesting to know whether there exists other coding schemes which require less prior knowledge, for example knowledge of $I_c := \chi_0 - \chi_1$ alone.

Our result on universal private coding can be generalized to the case of general quantum channels if some further information is available, as explained below. In the most general setting of transmitting private information over a quantum channel, $\mathcal{N}$, the sender Alice prepares a quantum input state ρ. Notice that there are unlimited number of ensemble

decompositions $\{p_i, \rho_i\}$ of ρ, satisfying $\rho = \sum_{i \in \mathcal{X}} p_i \rho_i$. Each such decomposition induces a $c \to qq$ channel, W^{BE}, from the isometric extension $U_\mathcal{N}$ of $\mathcal{N}$, such that the induced channels are given by $W^B : i \to \mathcal{N}(\rho_i)$ and $W^E : i \to \widehat{\mathcal{N}}(\rho_i)$, where $\widehat{\mathcal{N}}$ denotes the channel which is complementary to $\mathcal{N}$. By using Theorem 3, for each such ensemble $\{p_i, \rho_i\}_{i \in \mathcal{X}}$, we can then design a universal coding scheme which achieves a private transmission rate equal to $I_c(\{p_i, \rho_i\})$. Moreover, if prior knowledge of an ensemble which would lower bound the Holevo quantities of all possible $c \to q$ channels W^B, and upper bound the Holevo quantities for all possible $c \to q$ channels W^E, is available, then we can achieve the universal private coding for this general setting by direct application of Theorem 3. The private transmission rate achieved would then be given by

$$I_p := \min_{\{p_i, \rho_i\}} I_c(\{p_i, \rho_i\}).$$

ACKNOWLEDGMENTS

Appendix A: Proof of (16)

Let M_d be the linear space of all $d \times d$ complex matrices. Then $\widetilde{\mathcal{K}}_{\underline{q}} \subseteq \mathrm{Span}\{A^{\otimes n}|y^n\rangle : A \in M_d, |y^n\rangle \in \mathcal{K}_{\underline{q}}\}$. For any fixed $|y^n\rangle \in \mathcal{K}_{\underline{q}}$, let

$$\mathcal{K}_q(y^n) = \mathrm{Span}\{A^{\otimes n}|y^n\rangle : A \in M_d\}.$$

It then follows from Ref.[19] that

$$|\mathcal{K}_{\underline{q}}(y^n)| \leq (n+1)^{d^2}.$$

By dimension counting, we have

$$|\widetilde{\mathcal{K}}_{\underline{q}}| \leq \sum_{y^n \in \mathcal{T}_y^n(\underline{q})} |\mathcal{K}_{\underline{q}}(y^n)| \leq (n+1)^{d^2} |\mathcal{K}_{\underline{q}}|,$$

where $|\mathcal{T}_y^n(\underline{q})| = |\mathcal{K}_{\underline{q}}|$

Appendix B: Commutation relations

Lemma 7. *For any $\sigma \in \mathcal{D}(\mathcal{H})$, and τ_n defined by (17), the commutator $[\tau_n, \sigma^{\otimes n}] = 0$.*

Proof. Let $\mathcal{Y} = \{1, 2, \cdots, d\}$ and $d = \dim \mathcal{H}$. Recall that

$$\sigma^{\otimes n} = \sum_{\underline{q} \in \mathcal{P}_{\mathcal{Y}}^n} 2^{-n\left[D(\underline{q}\|\underline{\lambda}) + H(\underline{q})\right]} I_{\underline{q}},$$

where $\underline{\lambda} = (\lambda_1, \cdots, \lambda_d)$ is the vector of eigenvalues of σ and $I_{\underline{q}}$ is the projection operator defined by (14).

We have $I_{\underline{q}} \leq \widetilde{I}_{\underline{q}}$ from the fact that $\mathcal{K}_{\underline{q}} \subseteq \widetilde{\mathcal{K}}_{\underline{q}}$, $\forall \underline{q} \in \mathcal{P}_{\mathcal{Y}}^n$. It follows trivially that $[\tau_n, \sigma^{\otimes n}] = 0$. $\qquad\square$

Lemma 8. *Given a type $\underline{p} \in \mathcal{P}_{\mathcal{X}}^n$, and a sequence $x^n \in \mathcal{T}_{\mathcal{X}}^n(\underline{p})$, let x_o^n be the corresponding ordered sequence in $\mathcal{T}_{\mathcal{X}}^n(\underline{p})$. Then we have $[\omega_{x_o^n}, \tau_n] = 0$.*

Proof. By definitions of (17) and (22),

$$\omega_{x_o^n} = \left(\prod_{i=1}^k \frac{1}{|\mathcal{P}_{\mathcal{Y}}^{m_i}|}\right) \sum_{\underline{q}^{(\mathsf{x}_1)} \in \mathcal{P}_{\mathcal{Y}}^{m_1}} \cdots \sum_{\underline{q}^{(\mathsf{x}_k)} \in \mathcal{P}_{\mathcal{Y}}^{m_k}} \tau_{\underline{q}^{(\mathsf{x}_1)}} \otimes \cdots \otimes \tau_{\underline{q}^{(\mathsf{x}_k)}}.$$

It follows from (15) that the term $\tau_{\underline{q}^{(\mathsf{x}_1)}} \otimes \cdots \otimes \tau_{\underline{q}^{(\mathsf{x}_k)}}$ in the summand of $\omega_{x_o^n}$ can be written as:

$$\tau_{\underline{q}^{(\mathsf{x}_1)}} \otimes \cdots \otimes \tau_{\underline{q}^{(\mathsf{x}_k)}} = \frac{\widetilde{I}_{\underline{q}^{(\mathsf{x}_1)}} \otimes \cdots \otimes \widetilde{I}_{\underline{q}^{(\mathsf{x}_k)}}}{\prod_{i=1}^k |\mathcal{K}_{\underline{q}^{(\mathsf{x}_i)}}^{m_i})|},$$

where each $\underline{q}^{(\mathsf{x}_i)} = (q_1^{(\mathsf{x}_i)}, \cdots, q_d^{(\mathsf{x}_i)})$ is a type in $\mathcal{P}_{\mathcal{Y}}^{m_i}$. The following proof holds for every term of $\omega_{x_o^n}$. Define $I^* := I_{\underline{q}^{(\mathsf{x}_1)}} \otimes \cdots \otimes I_{\underline{q}^{(\mathsf{x}_k)}}$, where

$$I^* = \left(\sum_{y^{m_1} \in \mathcal{T}_{\mathcal{Y}}^{m_1}(\underline{q}^{(\mathsf{x}_1)})} |y^{m_1}\rangle\langle y^{m_1}|\right) \otimes \cdots \otimes \left(\sum_{y^{m_k} \in \mathcal{T}_{\mathcal{Y}}^{m_k}(\underline{q}^{(\mathsf{x}_k)})} |y^{m_k}\rangle\langle y^{m_k}|\right). \tag{B1}$$

For any sequence $y^n \in \mathcal{T}_{\mathcal{Y}}^{m_1}(\underline{q}^{(\mathsf{x}_1)}) \times \cdots \times \mathcal{T}_{\mathcal{Y}}^{m_k}(\underline{q}^{(\mathsf{x}_k)})$, the number of times any $i \in \mathcal{Y}$ appears in the sequence y^n is given by

$$N(i|y^n) = \sum_{j=1}^k m_j q_i^{(\mathsf{x}_j)},$$

where $q_i^{(\mathsf{x}_j)}$ is the ith element in the probability vector $\underline{q}^{(\mathsf{x}_j)}$. Such a sequence y^n must also belong to $\mathcal{T}_{\mathcal{Y}}^n(\underline{q})$ for the type $\underline{q} \in \mathcal{P}_{\mathcal{Y}}^n$, where the i-th element of $\underline{q}$ is

$$q_i = \frac{N(i|y^n)}{n} = \frac{\sum_{j=1}^k n p_j q_i^{(\mathsf{x}_j)}}{n} = \sum_{j=1}^k p_j q_i^{(\mathsf{x}_j)}.$$

27

In short, we can write $\underline{q} = \sum_{j=1}^{k} p_j \underline{q}^{(\varkappa_j)}$. Therefore $\mathcal{T}_{\mathcal{Y}}^{m_1}(\underline{q}^{(\varkappa_1)}) \times \cdots \times \mathcal{T}_{\mathcal{Y}}^{m_k}(\underline{q}^{(\varkappa_k)}) \subset \mathcal{T}_{\mathcal{Y}}^{n}(\underline{q})$, and

$$I^* \le I_{\underline{q}}.$$

Furthermore, we can obtain

$$\widetilde{I}^* \le \widetilde{I}_{\underline{q}}.$$

Therefore $[\tau_{\underline{q}^{(\varkappa_1)}} \otimes \cdots \otimes \tau_{\underline{q}^{(\varkappa_k)}}, \tau_{\underline{q}}] = 0$ for all $\underline{q} \in \mathcal{P}_{\mathcal{Y}}^{n}$ and for all $\underline{q}^{(\varkappa_i)} \in \mathcal{P}_{\mathcal{Y}}^{m_i}$. By linearity, we have $[\omega_{x_o^n}, \tau_n] = 0$. $\square$

Lemma 9. $[\omega_{x^n}, \tau_n] = 0$.

Proof. For any sequence $y^n \in \mathcal{T}_{\mathcal{Y}}^{m_1}(\underline{q}^{(\varkappa_1)}) \times \cdots \times \mathcal{T}_{\mathcal{Y}}^{m_k}(\underline{q}^{(\varkappa_k)})$, we know from Lemma 8 that $y^n \in \mathcal{T}_{\mathcal{Y}}^{n}(\underline{q})$, where $\underline{q} = \sum_{j=1}^{k} p_j \underline{q}^{(\varkappa_j)}$. Furthermore, sy^n must also belong to $\mathcal{T}_{\mathcal{Y}}^{n}(\underline{q})$, where $s \in S_n$ is a permutation such that $x^n = sx_o^n$. Denote

$$\mathcal{J}^n = \{sy^n : \forall y^n \in \mathcal{T}_{\mathcal{Y}}^{m_1}(\underline{q}^{(\varkappa_1)}) \times \cdots \times \mathcal{T}_{\mathcal{Y}}^{m_k}(\underline{q}^{(\varkappa_k)})\} \subset \mathcal{T}_{\mathcal{Y}}^{n}(\underline{q}).$$

Let $I_s^* = U_s I^* U_s^\dagger$, where I^* is defined in (B1). Obviously $I_s^* \le I_{\underline{q}}$. Then following the same argument as in Lemma 8, we can conclude that $[\omega_{x^n}, \tau_n] = 0$.

$\square$

REFERENCES

[1] A. S. Holevo, "The capacity of the quantum channel with general signal states," *IEEE Transactions on Information Theory*, vol. 44, pp. 269–273, 1998.

[2] B. Schumacher and M. D. Westmoreland, "Sending classical information via noisy quantum channels," *Physical Review A*, vol. 56, pp. 131–138, 1997.

[3] N. Cai, A. Winter, and R. W. Yeung, "Quantum privacy and quantum wiretap channels," *Problems of Information Transmission*, vol. 40, pp. 318–336, 2004.

[4] I. Devetak, "The private classical capacity and quantum capacity of a quantum channel," *IEEE Transactions on Information Theory*, vol. 51, pp. 44–55, 2005.

[5] S. Lloyd, "The capacity of a noisy quantum channel," *Physical Review A*, vol. 55, pp. 1613–1622, 1997.

[6] P. W. Shor, "The quantum channel capacity and coherent information," MSRI workshop on quantum computation, 2002. [Online]. Available: http://www.msri.org/publications/ln/msri/2002/quantumcrypto/shor/1/

[7] N. Datta and T. C. Dorlas, "Coding theorem for a class of quantum channels with long-term memory," *J. Phys. A: Math. Theor.*, vol. 40, pp. 8147–8164, 2007.

[8] I. Bjelaković and H. Boche, "Classical capacity of averaged quantum channels," *Proceedings of the 2008 IEEE International Symposium on Information Theory*, pp. 86–90, 2008.

[9] I. Bjelaković, H. Boche, and J. Nötzel, "Quantum capacity of a class of compound channels," *Physical Review A*, vol. 78, p. 042331, 2008.

[10] I. Bjelaković, H. Boche, and J. Nötzel, "Entanglement transmission and generation under channel uncertainty: Universal quantum channel coding," *Communications in Mathematical Physics*, vol. 292, no. 1, pp. 55–97, 2009.

[11] J. Wolfowitz, "Simultaneous channels," *Arch. Rational Mech. Anal.*, vol. 4, no. 4, pp. 371–386, 1960.

[12] J. Wolfowitz, Coding Theorems of Information Theory. Berlin: Springer-Verlag, 1978.

[13] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *Ann. Math. Stat.*, vol. 30, no. 4, pp. 1229–1241, 1959.

[14] M. Hayashi, "Universal coding for classical-quantum channel," *Communications in Mathematical Physics*, vol. 289, no. 3, pp. 1087–1098, 2009.

[15] I. Csiszár and J. Körner, Information Theory: coding theorems for discrete memoryless systems. New York-San Francisco-London: Academic Press, 1981.

[16] M.-H. Hsieh, I. Devetak, and A. Winter, "Entanglement-assisted capacity of quantum multiple-access channels," *IEEE Transactions on Information Theory*, vol. 54, no. 7, pp. 3078–3090, 2008.

[17] A. Winter, ""Extrinsic" and "intrinsic" data in quantum measurements: asymptotic convex decomposition of positive operator valued measures," *Communications in Mathematical Physics*, vol. 244, no. 1, pp. 157–185, 2004.

[18] M.-H. Hsieh, Z. Luo, and T. Brun, "Secret-key-assisted private classical communication capacity over quantum channels," *Physical Review A*, vol. 78, no. 4, p. 042306, 2008.

[19] R. Jozsa, M. Horodecki, P. Horodecki, and R. Horodecki, "Universal quantum information compression," *Physical Review Letters*, vol. 81, pp. 1714–1717, 1998.

[20] R. Jozsa and S. Presnell, "Universal quantum information compression," *Proc. Roy. Soc. London Ser. A*, vol. 459, pp. 3061–3077, 2003.

[21] M. Hayashi and K. Matsumoto, "Quantum universal variable-length source coding," *Physical Review A*, vol. 66, no. 2, p. 022311, 2002.

[22] M. Hayashi and K. Matsumoto, "Universal distortion-free entanglement concentration," *Physical Review A*, vol. 75, p. 062338, 2007.

[23] R. Blume-Kohout, S. Croke, and D. Gottesman, "Streaming universal distortion-free entanglement concentration," 2009, arXiv:0910.5952.

[24] F. Buscemi and N. Datta, "The quantum capacity of channels with arbitrarily correlated noise," *IEEE Transactions on Information Theory*, vol. 56, pp. 1447–1460, 2009.

[25] R. Koenig and S. Wehner, "A strong converse for classical channel coding using entangled inputs," *Physical Review Letters*, vol. 103, p. 070504, 2009.

[26] M. Ohya and D. Petz, Quantum entropy and Its Use. Berlin Heidelberg New York: Springer-Verlag, 1993.

[27] M. Mosonyi and F. Hiai, "On the quantum Rényi relative entropies and related capacity formulas," arXiv:0912.1286v4.

[28] N. Datta and R. Renner, "Smooth entropies and the quantum information spectrum," *IEEE Transactions on Information Theory*, vol. 55, no. 6, pp. 2807–2815, 2009.

[29] A. Winter, "Coding theorem and strong converse for quantum channels," *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 2481–2485, 1999.

[30] T. Ogawa and H. Nagaoka, "A new proof of the channel coding theorem via hypothesis testing in quantum information theory," *IEEE International Symposium on Information Theory (ISIT)*, p. 73, 2002, quant-ph/0208139.

[31] R. Ahlswede and A. Winter, "Strong converse for identification via quantum channels," *IEEE Transactions on Information Theory*, vol. 48, pp. 569–579, 2002.

[32] T. M. Cover and J. A. Thomas, Elements of Information Theory, ser. Series in Telecommunication. New York: John Wiley and Sons, 1991.

[33] M. A. Nielsen and I. L. Chuang, Quantum Computation and Quantum Information. New York: Cambridge University Press, 2000.

[34] M. Hayashi and H. Nagaoka, "General formulas for capacity of classicalquantum channels," *IEEE Transactions on Information Theory*, vol. 49, pp. 1753–1768, 2003.

[35] Z. Luo and I. Devetak, "Channel simulation with quantum side information," *IEEE Transactions on Information Theory*, vol. 55, pp. 1331–1342, 2009.

[36] I. Devetak, P. Hayden, and A. Winter, Principles of Quantum Information Theory, 2006, in preparation.

[37]G. Grimmett and D. Stirzaker, Probability and Random Processes. Oxford University Press, 2001.